



Машинное обучение для решения исследовательских и инженерных задач в науках о Земле

Михаил Криницкий

К.Т.Н., Н.С.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и
мониторинга климатических изменений (ЛВОАМКИ)

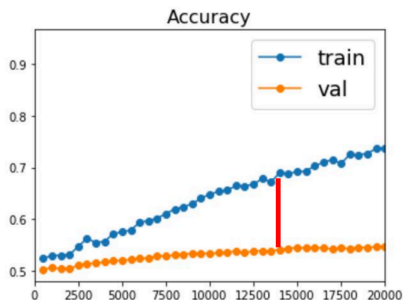
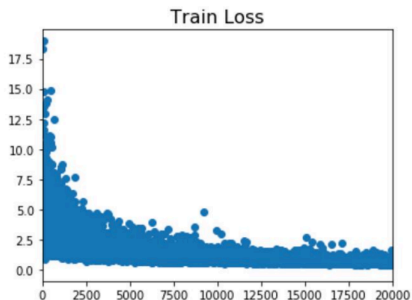
DEEP LEARNING for Earth Sciences

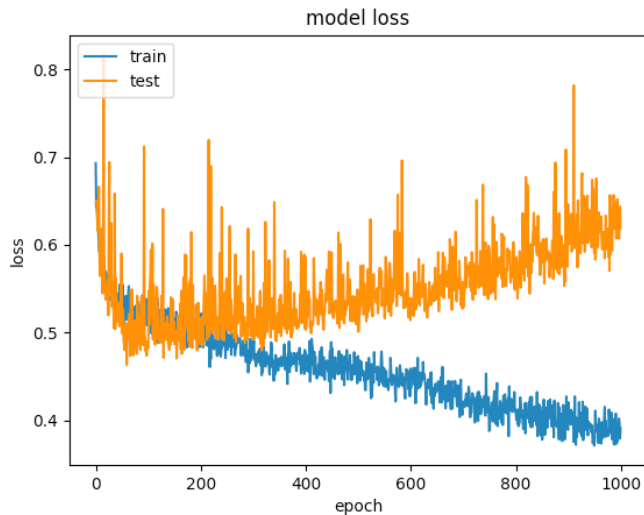
Святослав Елизаров

Институт океанологии им. П.П. Ширшова РАН

РЕГУЛЯРИЗАЦИЯ

Переобучение (overfitting) – явление, когда построенная модель хорошо работает на обучающей выборке и плохо на тестовой.





Идея

Давайте ограничивать норму весов.

$$\text{Loss} = \text{Data Loss} + \lambda (\text{Regularization Loss})$$

Для примера возьмем MSE функционал ошибки

$$L(x, w, y) = \sum_i (NN(x_i, w) - y_i)^2$$

Функционал с L1 регуляризацией

$$L(x, w, y) = \sum_i (NN(x_i, w) - y_i)^2 + \lambda \|w\|_{L_1}$$

Функционал с L2 регуляризацией

$$L(x, w, y) = \sum_i (NN(x_i, w) - y_i)^2 + \lambda \|w\|_{L_2}$$

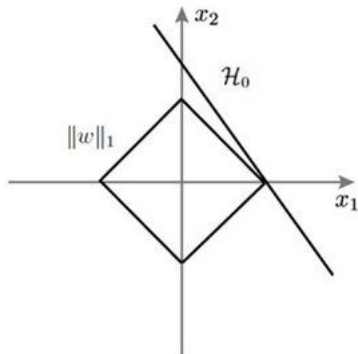
Функционал со смешанной регуляризацией

$$L(x, w, y) = \sum_i (NN(x_i, w) - y_i)^2 + \alpha \|w\|_{L_1} + \beta \|w\|_{L_2}$$

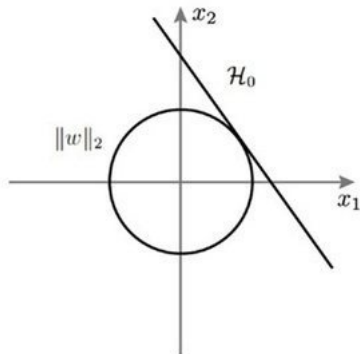
Везде α, β, γ — параметры, которые подбираются вручную.

L1 И L2 РЕГУЛЯРИЗАЦИЯ

L1 regularization



L2 regularization



Идея

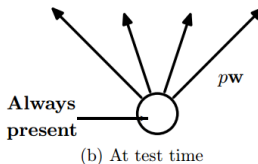
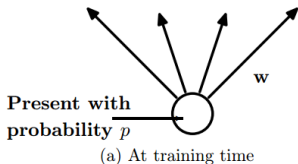
Давайте попробуем добавить шум в нейронную сеть для избежание переобучения

Во время обучения

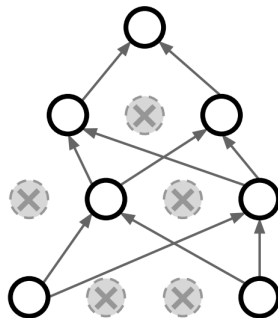
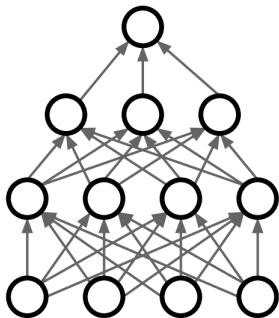
Для весов слоя W , создается маска M_W , в которой элементы с вероятностью p равны 1 и $(1-p)$ равны 0

Во время инференса

Мы просто умножаем наши веса на эти вероятности.



DROPOUT



- Значение вероятности p можно подбирать с помощью кросс-валидации, **(но не для больших задач)**
- Авторы утверждают, что p близкое к 0.5 хорошо подходит для большого количества сетей и задач
- При этом для инференса лучше использовать p близкое к 1 (или даже $p = 1$)

Больше информации:

Dropout: A Simple Way to Prevent Neural Networks from Overfitting
Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov; 15(Jun):1929–1958, 2014.

Идея dropout

Давайте попробуем добавить шум в нейронную сеть для избежания переобучения

Идея variational dropout

А давайте теперь ещё попробуем выучить каким должен быть этот шум

Пусть A – матрица входных данных

$$y_i = (x_i \cdot \xi_i) \theta_i, \quad \xi_{i,j} \sim N(1, \alpha)$$

Это можно переписать в виде:

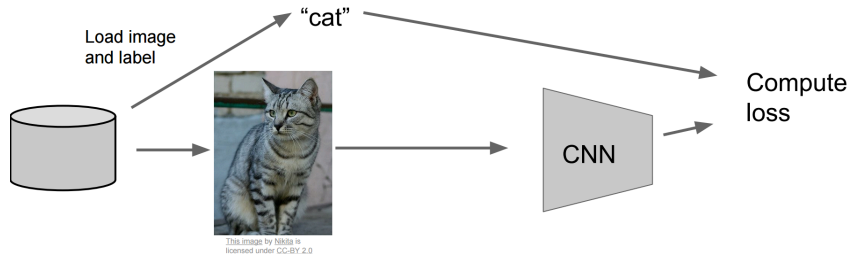
$$w_i = \xi_i \cdot \theta_i; \quad y_i = x_i w_i$$

Обучается с помощью вариационного инференса и с применением reparametrization trick. (Про это будет отдельно)

Больше информации:

Variational Dropout and the Local Reparameterization Trick
Diederik P. Kingma, Tim Salimans and Max Welling
<https://arxiv.org/abs/1506.02557>

DATA AUGMENTATION



DATA AUGMENTATION

